Design and implementation of an innovative dichotomous classification e-key to identify fish species

Emmanouil Tolis
Department of Informatics,
Alexander TEI of Thessaloniki,
P.O. Box 141, 57400,
Thessaloniki, Greece
E-mail: emtolis@it.teithe.gr

George Minos
Department of Aquaculture & Fisheries,
Alexander TEI of Thessaloniki,
P.O. Box 157, 63200,
N. Moudania, Greece
E-mail: gminos@aqua.teithe.gr

Vassilis Kostoglou*
Department of Informatics,
Alexander TEI of Thessaloniki
P.O. Box 141, 57400,
Thessaloniki, Greece
E-mail: vkostogl@it.teithe.gr

Abstract

A dichotomous identification key organizes fishes based on their similarities and differences, and is an essential tool which ichthyologists, biologists, fishery biologists and others can use for accurate identification of specimens or finding information on correct names, biology and distribution of species. This research work focuses on the development and implementation of a totally new information system which is able to identify correctly fish species names. The developed system is a fully interactive fish identification e-key which can be used in both forms; locally and remotely via the World Wide Web. This new dichotomous classification e-key provides the capability to identify any species in a compact and easy to use environment which gives the user excellent operation capabilities and complete information about all included fish species.

1. INTRODUCTION

There are about 28,000 living species of fishes which makes very difficult their correct classification (Helfman et al., 2009). The identification of the various fish species is based on morphometric characters (measurable structures such as fin length, head length, etc.), meristic characters (countable structures such as number of scales in the lateral line, number of vertebrae, etc.), anatomical characters of the skeleton and the soft anatomy or characters than include any fixed, describable differences among taxa such as color (presence of stripes, spots) photophores (number and position) and sexually dimorphic structures (Bristow 1992; Helfman et al., 2009).

To classify different species a dichotomous identification key is used, which is an extremely important tool in science. The primary aims of an identification key are to enable species to be identified correctly and to summarize what is known on their biology and geographical distribution. In order to identify a fish with the use of a dichotomous key, the user works through a series of questions and illustrations which eventually lead him to the species matching best the characteristics he has set.

Systematics (or taxonomy) is the biological science responsible for the classification of living organisms in a hierarchically organized system representing the evolutionary kinship of the various systematic groups. In classification, use of morphological, anatomical, physiological and other characteristics is made to decide the existing relationships (Bristow, 1992). The basic systematic unit (taxon) is the species followed (in ascending order) by the genus, family, order, class, superclass, subphylum and the phylum.

Like other animals and plants, fishes are known by common names and scientific names. While common names differ from country to country, scientific names are universal. Aristotle was the first to classify the animals known in his days, but the first generally acknowledged scientific classification of animals and plants was by Carl Linnaeus who introduced the binomial system, in which every species was given two Latin or Greek names. Since the scientific name consists of two parts, the first italicized word, with the initial letter capitalized, is the genus while the second italicized word is the specific (species) name.

The existing identification keys are divided into two categories: 1) printed keys and 2) electronic keys (e-keys). The former are printed in the form of a book and they mostly still keep this form so far. Nevertheless, e-keys have been developed in recent years. Examples of printed identification keys are books dealing with fishes from Greek seas (Kaspiris, 2000), Mediterranean Sea (Whitehead et al., 1984; Fischer et al., 1987*)*, Atlantic Ocean (Perlmutter, 1961; Fischer et al., 1981; Carpenter, 2002), Indian & Pacific Ocean (Fischer & Whitehead, 1974; Carpenter & Niem, 1998-2001; Fischer et al., 1984). The main disadvantage of the printed identification keys is that it is easy to make a few wrong decisions when navigating through the test. So, when someone is deadlocked or makes a wrong selection, it is not easy to go to a previous selection (family, suborder etc.) because there is not an area that shows the history of the selections. In a case like this, the reader has to find the previous selections that he made and the page in which they were. Furthermore, when the selections are too many, it leads to confusion for the reader. Also, it must be mentioned the case where new dichotomous keys must be created and printed which will be used to identify new organisms (fish species) that will appear in the area in the future.

For this reason, fish identification keys evolved as e-keys. There are many examples of fish identification e-keys because, as technology boomed, several keys of this type were developed to facilitate ichthyologists, students or persons that needed such tools. All fish identification e-keys are based on printed keys (books). No new keys have been created, but the book contents have been digitized. Some examples of identification keys in World Wide Web are in Fish Base (Froese and Pauly, 2011) per FAO area, per order or per family or quick identification by image and also identification by morphometrics. An important tool is the Marine Species Identification Portal (http://species-identification.org/) while the Fish Identification Site (http://svrsh2.kahaku.go.jp/fishis/) helps to identify fishes utilizing countable characters such as numbers of fin rays, scales, pores, gill rakers, body rings, and vertebrae. The common feature of these keys is that they are web applications. This means that they are uploaded as webpages into a website and when someone wants to use them, he just has to visit the specific website.

A dichotomous key is a tool that allows the user to determine the identity (specific name) of a fish. These keys consist of a series of "either or" choices that lead the user to the correct name. "Dichotomous" means "divided into two parts". Therefore, dichotomous keys always give two choices in each step.

Technically, there are two types of e-keys. The more simple e-keys (with static content) are developed with HTML and contain a set of information which is stored and divided into a number of pages. These pages compose a webpage. The above set of information is about key's selections, data on fish species, fish images and all the necessary information which compose a fish identification key. A simple e-key is not flexible because it does not provide updating capabilities (with which the user can add new fish species). Furthermore, a simple key lacks proper organization because it does not contain any database which can provide organized information storage. The more complex e-keys are developed under both HTML and a scripting language. The scripting language is usually either PHP (Hypertext Pre Processor) or ASP (Active Server Pages). By using a scripting language the developer is able to create a webpage with dynamic content, ensuring also that all the necessary for the operation of the e-key information is being stored into a database. This e-key does not simply show to the user a set of information which is divided into a number of webpages, but every time the user makes a selection, a set of information is being recovered from the database to be shown to the browser. As e-keys of this type are using databases, they could provide updating capabilities. An extensive literature review of the existing fish identification e-keys revealed that there are no e-keys providing complete and correct update capability.

The information system that has been developed in the present work, constitutes the first Greek fish identification e-key for all the Mediterranean fish species. The designed system is fully interactive with the user and can be used in both forms; locally and remotely via the World Wide Web. As the information system constitutes a desktop application, it provides an easy and user-friendly environment which gives the user multifaceted fish identification capabilities and a fast search function for all included fish species. Furthermore, its navigation function is a strong and useful advantage. Finally, its additional function which shows information about the fish systematic taxonomy is innovative.

2. METHODOLOGY

The information system has been developed with the Java object-oriented programming language. The system was necessary to have a database to connect to, so that to recover from this all necessary data and also to present this data as information to the end user. The database includes full information about every level of the fish systematic taxonomy, as well as images of every fish species. The RDBMS that used to manage the database of the information system is MySQL. Its choice made so that to provide to the information system the capability to connect to a single database, which is common to all users as it is uploaded to a Web Server. The above feature is something usual for MySQL and it is not found in all other RDBMS's, either they are free or not. Of course, MySQL can also be used only locally at the personal computer of a user. In this case, each user's personal computer takes Web Server's role. So, each personal computer keeps the database stored in it.

Figure 1 presents the structure of the database through an ER (Entity-Relationship) diagram including the main table of the database containing information about all selections of the user. All other tables have direct link with this table, because every level of the fish systematic taxonomy hierarchy participates in the fish identification key selections.
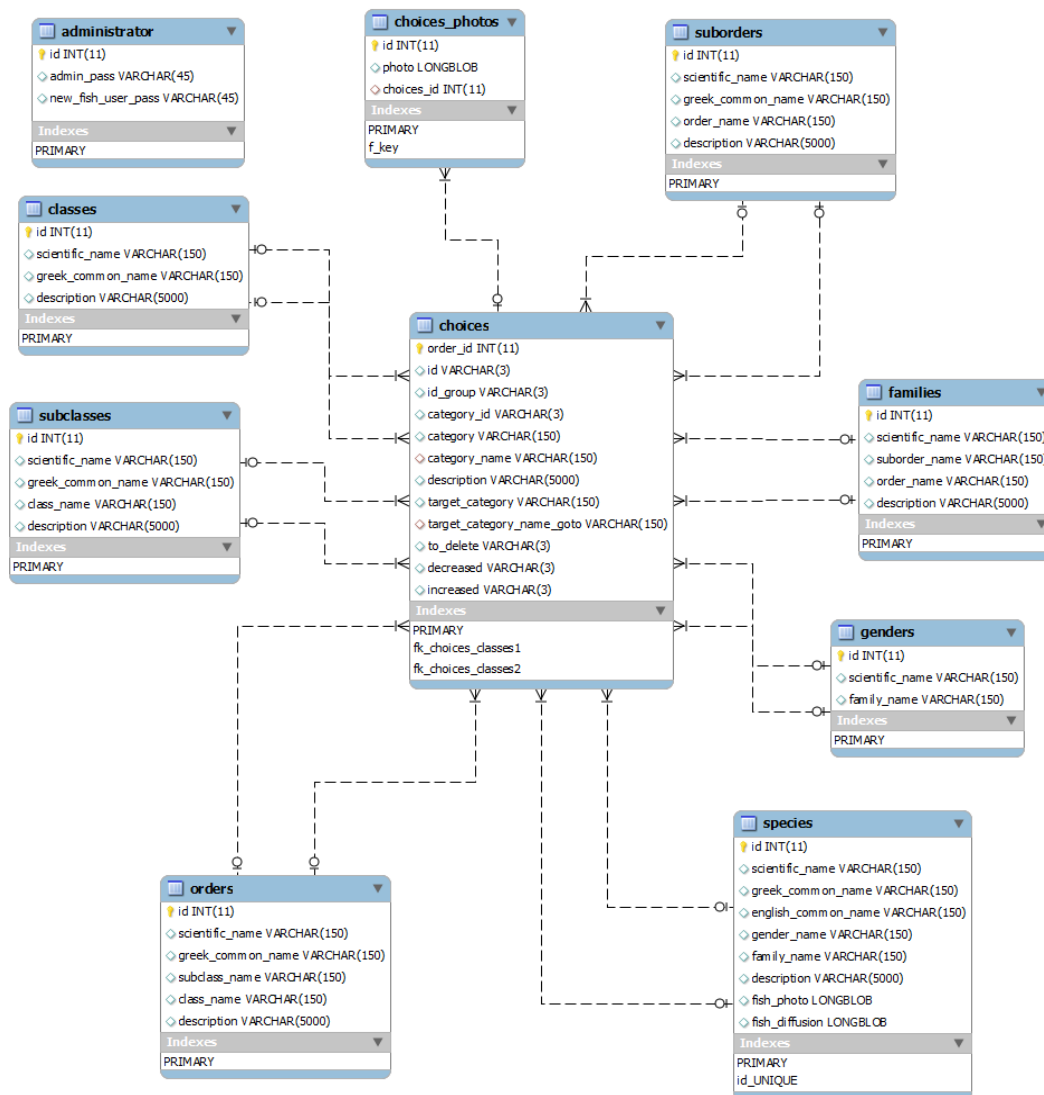

Figure 1. Information system's database structure through an ER diagram

Two programming tools have been used for the development of the information system: Neetbeans IDE 6.9.1 for the programming part, and MySQL Workbench 5.2.31 CE for database design and management. Apart of the above basic tools, a Java library was also used. This library is the Ganymed SSH-2 for Java library and was used so that to support the development of the information system. This library implements the SSH-2 network protocol and gives to a Java program the capability to connect with an SSH Server.

3. THE DEVELOPED INFORMATION SYSTEM

The information system provides the user the capability to identify a fish species by making some selections relevant to species' external morphology. When starting the application the user has to choose one of the two databases (local or online) and type a password. After that, the main menu of the e-key appears at the upper part of the screen (Figure 2).

The screen is divided into three horizontal parts. The upper part and the bottom part are static. Conversely, the central part is dynamic. In the upper part of the screen there are application's three main functions: 1) the "Show all species" function, 2) the "Search species" function, and 3) the "Insert new species" function. The bottom part constitutes a status bar which contains information about the connected database and a bar which informs the user about the progress of a specific search procedure.
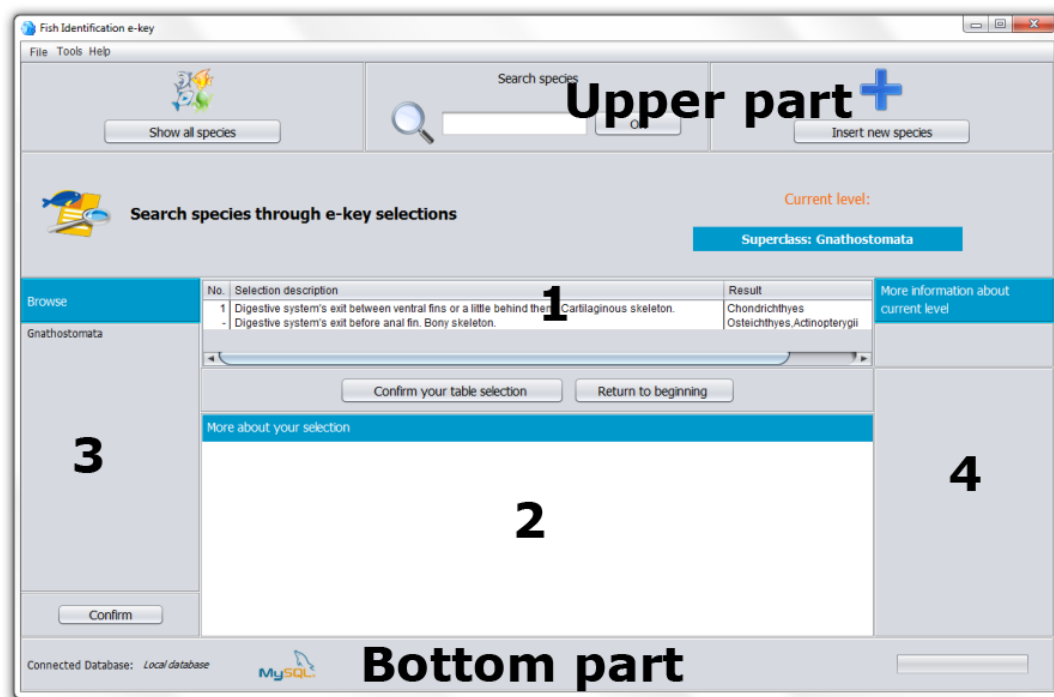

Figure 2. Main screen of the fish identification e-key

The central part of the e-key (Figure 2) consists of the following sections: 1) the table at the centre which contains the texts of the current dichotomous selection and two buttons, one to confirm the selection and another to return to the beginning, 2) the area below the table which shows more information to help user's selection, 3) the navigation at the left side and 4) the area at the right side which shows more information about the current level in which user is based on his selections. The upper section of the centre part (above the table) contains a title which informs the user about the application's function that is being executed at that time.

If the user intends to use the fish identification function, he has to select one of the two selections in the central part (area 1) and press the button "Confirm your table selection". Anytime he wants to return to the selections in the start of the key, he can just press the button "Return to beginning". Every time the user

makes a selection, he can read its full description (area 2). Also, when it is necessary, this area shows an image which describes optically user's selection. Furthermore, every user's selection is being recorded in the navigation list (area 3), so that he can anytime go back to a previous step. This list is very useful, as it presents the whole route until the final successful fish identification.

The left selection of the upper part (area 1 – "Show all species") leads to a table including all the fish species that are stored at the selected database until that time. This area also contains two buttons ("More information about the species" and "Back to e-key"). The former requires the selection of a row from the fishes' table and presents all the stored information about the selected species including names, picture, description and geographical distribution (Figure 3). The latter button restores the fish identification screen at the centre part of the e-key.
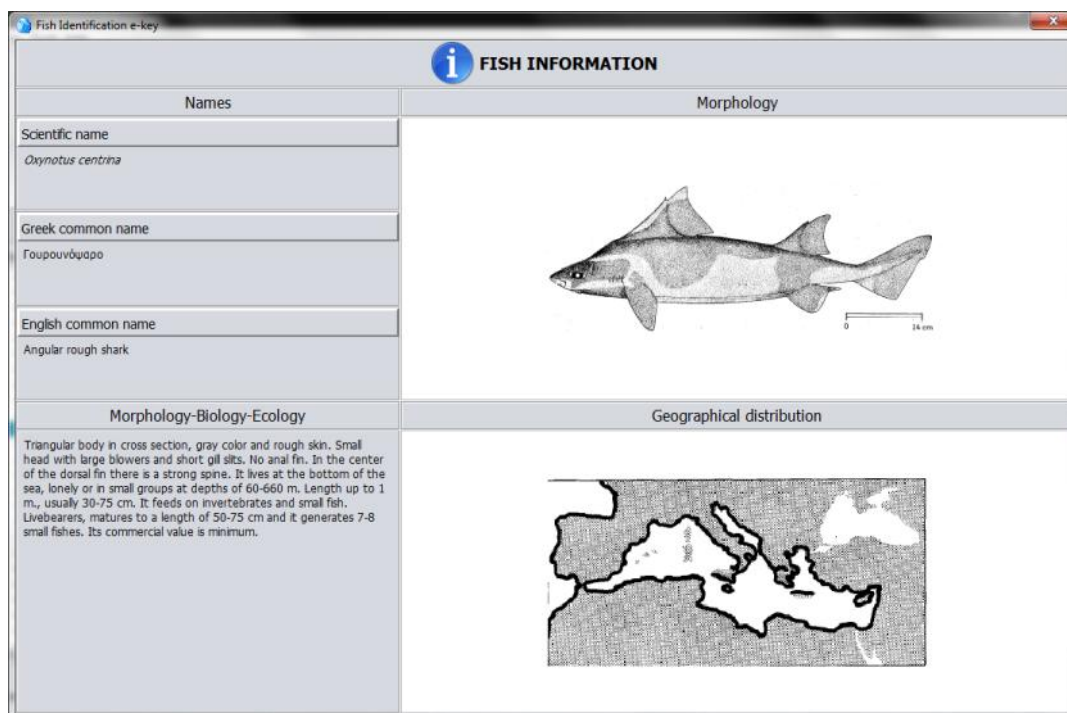
Figure 3. Fish identification – Window with species information

In Figure 4 appears a snapshot from the application presenting a full catalogue of all the fish species which are stored to the connected database and including their scientific names, Greek common names and English common names.

4. CONCLUSIONS AND FURTHER RESEARCH

The developed information system constitutes the most modern and functional fish identification e-key, compared with the existing ones. Both its additional functions and its innovation make it special. It provides to the user multiple search and fish identification capabilities. The fish species search can be performed either by searching by the species name, or by applying the selection procedure through dichotomous questions. In this procedure the user reads the selection's description and sees a photo, so that he can make the selection that matches more the external features of the species he is looking for. Furthermore, the information system is a desktop application which can be installed on any personal computer. This feature makes it functionally faster than other respective web applications. Also, its search capability gives very fast results, as well as an organized and detailed presentation of fish species information. The navigation function, which is being enabled every time a user is trying to identify a species, is a strong and useful advantage. Finally, is worth mentioning the additional function which presents information about the fish systematic taxonomy.

Figure 4. Catalogue with all the fish species included in the database of the e-key

Regarding forthcoming activities, firstly any disadvantages or errors that will be identified during the use of the information system have to be improved and debugged respectively. Some issues for further relevant research are the following: 1) An additional function to the application, which through a web connection will provide access to a developed forum to discuss new species' additions, comments on the application etc., 2) Application's conversion to a mobile application, for use by mobile phones, smartphones or tablets with touchscreen, 3) Application's extension to a wider geographical area by adding more fish species, and 4) The most innovative new feature would be the addition of updating capabilities which will allow users to add to the existing database new fish species. Current research of the authors of the present article focuses on the requirement analysis and implementation of this specific issue.

REFERENCES

Bristow, P. 1992. *The illustrated encyclopedia of fishes*. Chancellor Press, London.

Carpenter K.E. 2002. The living marine resources of the Western Central Atlantic. Vol. 1-3. *FAO Species Identification Guide for Fishery Purposes and American Society of Ichthyologists and Herpetologists,* Special Publication No. 5. Rome, FAO. 2002. pp. 1-2127.

Carpenter, K.E. and V.H. Niem, 1998-2001. *FAO species identification guide for fishery purposes. The living marine resources of the Western Central Pacific*. Vol. 2-6. Rome, FAO, pp. 678-4218.

Fischer W. and Whitehead P.J.P. 1974. *FAO species identification sheets for fishery purposes. Eastern Indian Ocean (fishing area 57) and Western Central Pacific (fishing area 71)*. Rome, FAO, Vol. 1-4: pag.var.

Fischer W., Bianchi G. and Scott W.B. 1981. *FAO species identification sheets for fishery purposes. Eastern Central Atlantic. Fishing Area 34 and part of 47.* Canada Funds-in-Trust. Ottawa. Department of Fisheries and Oceans Canada, by arrangement with the Food and Agriculture Organisation of the United Nation. Vols. 1-6: pag. var.

Fischer W. and Bianchi G. 1984. *FAO species identification guide for fishery purposes. Western Indian Ocean (Fishing Area 51).* Prepared and printed with the support of the Danish International Development Agency (DANIDA). Rome, Food and Agricultural Organization of the United Nations, Vols 1-6: pag. var.

Fischer W., Bauchot M.L. and Schneider M. 1987. *Fiches FAO d'identification des espèces pour les besoins de la pêche. (Révision 1). Méditerranée et mer Noire. Zone de pêche 37*. Vol. II. Vertébrés. Publication préparée par la FAO, résultat d'un accord entre la FAO et la Commission des Communautés Européennes (Projet GCP/INT/422/EEC) financée conjointement par ces deux organisations. Rome, FAO, Vo1.2: 761- 1530

Froese R. and Pauly D. 2011. *FishBase. World Wide Web electronic publication.* www.fishbase.org, version 02/2011.

Helfman G.S., Collette B.B., Facey, D.E. and Bowen, B.W. 2009. *The diversity of fishes. Biology, evolution, and ecology.* Second edition. Wiley-Blackwell, UK.

Kaspiris P., 2000. *The Fishes of Greece* (*Identification Keys).* TYPOffset Manoudi-Stanidi, Patras, Greece (In Greek).

Minos G., 2008. *Fish Biology and Systematics*. Second Part. ATEI of Thessaloniki, Department of Aquaculture & Fisheries Technology, Greece. (In Greek).

Perlmutter, A. 1961. *Guide to Marine Fishes*. Bramhall House, New York.

Whitehead P.J.P., Bauchot M.L., Hureau J.C., Neilsen J. and Tortonese E. (eds.) 1984. *Fishes of the North-eastern Atlantic and the Mediterranean*. Vol. I-III, Unesco, Paris.